A Control Theory of Action*

Mikayla Kelley

University of Chicago

Penultimate draft. Forthcoming in *Australasian Journal of Philosophy*

Abstract

One of the central problems in the philosophy of action is to spell out the distinction between action and what merely happens, e.g., a wink versus an eye twitch. This essay proposes a theory of action offering an account of this distinction. The central claim of the theory is that action is movement that is controlled by the mover, where movement is understood capaciously and control is characterized by a trio of conditions consisting of an aim condition, a modal condition, and an explanatory condition. Importantly, being controlled is shown to be a determinable property of movements, and its determination dimensions are isolated. Paired with the claim that action is movement that is controlled by the mover, this account of control as a determinable illuminates the diversity within the category of action.

1. Introduction

Consider the difference between a wink and an eye twitch, raising an arm and an arm spasm, a squirrel jumping off a tree and the squirrel being pushed off the tree, flipping one's hair while dancing and flipping one's hair while having a seizure, or imagining a sunrise and the image of a sunrise popping to mind. There is an intuitive distinction drawn in these pairs of examples. This essay tries to illuminate this distinction by offering a solution to *the problem of action*, that is, the problem of explicating the difference between an individual's actions and what merely happens to the individual.[1]

The orthodox solution to the problem of action is Donald Davidson's causal theory of action on which actions are bodily movements caused in the right way by belief-desire pairs (Davidson 1980*a*). Focusing instead on the

*For many discussions and rounds of feedback, thanks to Michael Bratman and Antonia Peacocke. For helpful comments, thanks to Dan Friedman, Wes Holliday, Will Hornett, Kiran Luecke, Taylor Madigan, Joseph Martinez, Austen McDougal, Juan Piñeros Glasscock, Josh Shepherd, and Elise Sugarman. Thanks also to audiences at Stanford University, University of Chicago, University of Pennsylvania, New York University, University of Pittsburgh, University of Southern California, Australian National University, Georgia State University, and the Non-Human Agency and the Standard Story of Action Workshop. Finally, thanks to multiple anonymous referees for their careful and incisive comments.

[1] In Frankfurt's (1978) words: 'the problem of action is to explicate the contrast between what an agent does and what merely happens to him, or between the bodily movements that he makes and those that occur without his making them' (p. 157).

way that agency often involves significant cross-temporal organization, we might think about action in terms of causal sustenance by intentions understood as plan states (Bratman 1987). Focusing on the particular features of action at the time of acting (as opposed to its causal history), we might see action in terms of guidance by the agent, where this requires a compensatory mechanism within which the action is embedded (Frankfurt 1978). Focusing on the distinctive epistemological features of action, we might see action as the manifestation of practical knowledge or awareness of what one is up to when acting (Anscombe 1957). Or we might explain the distinction between action and mere happening by positing the existence of irreducible goal-directedness (Wilson 1989), irreducible causation by agents (Chisholm 1964), or two-way powers (Steward 2012).

I propose that control is the distinguishing feature of action. More precisely, action is movement that is controlled by the mover.[2] I will also show how control and thus action are *determinables*—control and thus action come in more specific forms much in the way that color, taste, and shape come in more specific forms. In seeing action as a determinable, my proposal will thereby offer a unification of all of these alternative proposals.

The plan for the rest of the essay is as follows. In Section 2, I discuss an important clarification of this essay's goal. Section 3 lays out the proposed solution to the problem of action, which I call the *control theory of action*. In Section 4, I apply the theory to some test cases to further clarify the theory's extensional commitments. In Section 5, I consider three objections to the purportedly tight connection between action and control on the control theory of action. I conclude in Section 6.

2. Clarifying the Goal

There are many phenomena associated with the category of action. At the broadest level, there is action in the sense of behavior or what is done. In this sense, anytime something does something or is engaged in a bit of behavior, there is an action. Another broad sense of action is that of causing a change (Alvarez and Hyman 1998; Hyman 2015). In this sense, the acid acts on the metal in eroding it and the sun acts on a piece of chocolate in melting it. Another broad sense of action includes the pre-representational behaviors of paramecia, ticks, and hydras (Burge 2009). On this conception, it is the meeting of some standard set by one's biological needs that marks the difference between action and what merely happens to an individual. Much of contemporary philosophy of action has focused on action in narrower senses than these. Many existing theories of action are theories of intentional action, for example. There is also interest in autonomous action, voluntary action, free action, and full-blooded human action. The narrower focus is motivated by widespread interest in

[2] As I will discuss in the next section, I actually establish the weaker claim that there is an important kind of action—what I call *reflexive action*—which is movement that is controlled by the mover. In work in progress, I argue that all action is reflexive. Thus, taken together, these two arguments establish the strong claim that action (as such) is movement that is controlled by the mover.

understanding action in senses that matter for ethics and understanding human agency in particular.

The control theory of action is a theory of a thus far neglected phenomenon associated with the category of action, a phenomenon which is narrower than action in the broad senses that concern Alvarez, Hyman, and Burge but broader than the narrow senses that concern most contemporary philosophers of action. It follows that the control theory of action is not in competition with many existing theories of action (though see Footnote 4, Footnote 36, and the end of Section 3). Instead, the theory illuminates a new and interesting sense of action. This sense can be gestured at with the idea of *action which constitutes a success by the agent's own lights*. Such actions involve *reflexivity*, that is, they meet some standard set by the agent whose action it is. Call this neglected agential phenomenon *reflexive action*.[3] The minimal conceptions of action that I mentioned above do not entail reflexivity. The primitive form of agency that Burge is concerned with involves the meeting of a standard but not (necessarily) one set by the agent. Causation of a change clearly does not entail any sort of success by the lights of the agent. The acid does not in any sense succeed by its own lights when it erodes the metal—in fact, the acid has no lights at all. In the other direction, as will become clear as I explicate the idea of reflexivity, intentionality is only one form of reflexivity. Thus, action involving reflexivity is a broader class than action involving intentionality.

The control theory of action is a theory of reflexive action. Thus, a first desideratum on the theory is that its extension be roughly in line with the class of events we would intuitively describe as action involving a success by the lights of the agent. A second desideratum is that the theory have the usual features of a good philosophical theory—it is, for example, simple, fecund, illuminating, and joint-carving. A third desideratum is that the theory show why reflexive action is a sense of action worth studying. Given that reflexive action hasn't been studied much, if ever, one might reasonably wonder whether this is for good reason. I won't have space to illustrate all the ways that reflexive action is theoretically, practically, and ethically important. But I will develop one particular advantage of studying reflexive action: we come to understand a sense of action that can unify many existing theories of ethically important forms of agency like intentional, autonomous, and free agency.

Another crucial motivation I have for studying reflexive action is that I think it is the broadest *genuine* form of action. All genuine action is reflexive, despite what Alvarez, Hyman, and Burge have said in defense of the

[3]This sense of action could also be gestured at with the phrase *representational agency*. Burge says that in representational agency, 'the agent does not act blindly. Its goals are represented, and set internally, by itself. The *agent* perceives, and acts to fulfill goals that it represents' (p. 277). See also Shepherd's (2021) discussion of *psychological agency* (Sec. 6.3). One might be tempted to gesture at this sense of action with the phrase *goal-directed agency*, but I think this would be a mistake. This phrase is suggestive of a sense of action that is both too strong and too weak to be reflexive action. The notion of goal-directedness is too strong, because goal-directedness suggests the existence of a feedback mechanism that is not required for reflexive action (see Section 3.3); the notion of goal-directedness is also too weak because something can be understood as goal-directed while involving no psychological or representational states—and thus while not constituting a success by the agent's own lights (e.g., the sunflower growing toward the sun).

acid and the hydra being capable of genuine action. So the control theory is not just a theory of reflexive action but a theory of action *in the broadest sense*. However, I cannot fully establish this strong thesis in this essay, since it requires attending to difficult methodological issues about what makes a theory of action successful as a theory of action in the broadest sense. Instead, I take only the first step of illuminating what exactly reflexive action is. Here, I will develop and defend the control theory of action as a theory of reflexive action, leaving an in-depth discussion of action in the broadest sense for a later time.[4] Suggestively and for ease of exposition, I will use 'action' to refer to reflexive action from now on; but the qualification discussed in this section should remain in view.

3. The Proposal

Even once we have identified our target phenomenon with the demarcating idea of action that constitutes a success by one's own lights, we are far from a theory of that phenomenon. We have only focused our gaze on a phenomenon about which a theory could be given. Putting forth this theory is the task for the rest of the paper.

In this section, I lay out the basic commitments of the control theory of action in four parts. In the first part, I specify the kind of thing action is, namely movement. In the second part, I reflect on the diversity within the category of action. In the third part, I offer a theory of the property that distinguishes the actions within the class of all movements, namely reflexive control. In the fourth part, I explain how the control theory of action offers a unification of many existing theories of action.

3.1. Action as Movement

A first question one meets in an inquiry into the metaphysics of action is: which ontological category do actions fall under? Do actions fall under a familiar category (such as events, processes, or causings of events) or must

[4] In fact, it's not so easy to set aside action in the broadest sense, since the class of reflexive actions falls within the class of actions in the broadest sense. So if the conditions required for action in the broadest sense are not met by my theory of reflexive action, I couldn't possibly have the right theory of reflexive action. I grant this point, but I don't know how to respond to it adequately without arguing that the control theory is the best theory of action in the broadest sense, which I can't do here. So let me just say a few things about how the control theory can be made consistent with or can improve upon existing theories of action in the broadest sense. First, if one thinks acting is causing a change, one could alter the control theory (which claims, roughly, that reflexive action is movement controlled by the mover) to the view that reflexive action is a causing of a change that is reflexively controlled by the causer. Second, according to Steward (2012), any action involves the exercise of a two-way power. Thus, it would seem to follow on this theory that if determinism is true, there are no actions (for discussion, see Alvarez 2013, Sec. 4). I likely cannot accommodate this strong of a view, but I can accommodate the weaker view that the exercise of a two-way power isolates an important kind of (reflexive) action (see Section 3.4 for how that might go). On this weaker view, determinism does not entail the existence of no actions at all, just no actions of a certain sort. Third, on Davidson's causal theory of action, all action is intentional under some description. As will become clear, intentionality is just one form of reflexivity, so reflexive action is broader than intentional action. A recurring objection to Davidson's causal theory of action is that it is too narrow as a theory of action in the broadest sense; thus, the control theory, while sharing some structural features with the causal theory, can improve upon it in being more permissive.

we introduce a new category to make room for them? I will build the control theory of action upon particular answers to these questions for the sake of concreteness, but the basic ideas of the theory should be consistent with different foundations.

As for the chosen foundation, the control theory of action follows a number of prominent theories of action—including event-causal theories of action (Aguilar and Buckareff 2010)—in that it identifies token actions with token events. I will make the further orthodox assumption that one's actions fall within a particular subclass of events, namely the subclass of one's bodily movements, doings, or behaviors. In taking actions to be within this particular subclass of events, one responds to a basic conceptual datum about action, namely that an action of mine must be an event that in some sense involves me as the subject. It is for this reason that your laughing, the London Bridge's falling, the tornado's destroying the town, and my dog's barking cannot, as a matter of conceptual fact, be my actions no matter the influence I have over them. These are simply the wrong kinds of events to be actions of mine. In other words, where we let the *extended action thesis* be the view that the actions of an individual can constitutively extend beyond the individual into the behaviors or movements of other individuals, I am claiming that the extended action thesis is false.[5] With the choice of terminology, I am purposefully drawing a parallel to the extended mind thesis (Clark and Chalmers 1998). I see the extended action thesis as an analogue of the extended mind thesis; and just as the extended mind thesis is often admitted to be false about our actual concept of mind, the extended action thesis is plausibly false about our actual concept of action.[6]

Now, this important subclass of events—those that involve the individual as, in some sense, the subject—is only crudely gestured at with imprecise phrases like 'bodily movement' and 'behavior'. Here, I suggest a more precise account: the actions of an individual fall within the subclass of events that consist in changes in the mental and/or physical properties of the individual.[7] To link my view with the predominantly held *Corporealist* view according to which actions are bodily movement events, I will refer to such events as *movements of that individual*. Thus, we have an answer to the ontological question with which we began: actions are movements.[8]

[5]For simplicity, I will assume throughout that an individual is exhausted by their mind and body. Likely, technological advances will undermine this picture of individuals, especially in light of recent developments in virtual reality and brain-computer interfaces. But I leave adjusting the control theory of action accordingly for another time. In any case, I hope it becomes clear that the control theory of action illustrates a point made persuasively by O'Brien (forthcoming): the question of what an action is is intimately bound up with the question of what an individual is.

[6]See Ford 2018 for a defense of the extended action thesis. While I take the falsity of the extended action thesis to be a commonsense feature of the way we think about action and so will assume as much, this assumption could be easily extracted to form a control theory of *extended* action (see Footnote 34). Thus, there is also a sense in which the control theory of action is neutral with respect to the extended action thesis; and so if one is convinced by much of what I come to say but not the falsity of the extended action thesis, one is free to develop a control theory of extended action instead.

[7]Of course, the descriptions of our actions often refer to outside the body and mind. Our actions include building houses, writing letters, baking cakes, and starting wars. The thought behind the identification of token actions with token changes in the mental and/or physical properties of an agent is that tokening one of these act types *consists* in such changes. Davidson's 'accordion effect' is relevant to this point; see also Bratman 2006.

[8]One important objection to event-based theories of action is that they purportedly cannot account for omissions as actions

3.2. The Generality of Action

In the rest of this section, we turn to what I see as the core contribution of the control theory of action. Clearly not all movements are actions. What property distinguishes the actions within the class of all movements?

Recall that in Section 2, I narrowed our focus to a particular sense of action, one which is broader than intentional action and other strong senses of agency that have been the focus of much of contemporary philosophy of action. Thus, as should be expected, many properties of interest in contemporary philosophy of action are at the wrong level of generality to be this property which differentiates between the (reflexive) actions and the mere movements. For one, a number of different phenomena can set a standard on one's behavior, including motivating reasons, intentions, normative judgments, desires, motor representations, and pieces of practical knowledge. A difference in which of these phenomena is the origin of a movement may entail a difference in intentionality, but intuitively they all suffice for *some* form of action.[9] Similarly, the explanatory structures underlying action can differ, from action to action, creature to creature, and possible world to possible world. One's aim might causally explain one's action, as the orthodox causal theory of action emphasizes, but the explanation need not be causal: the aim might constitutively explain one's action, as when one aims to please God and thereby, in simply forming this aim, pleases god.[10] While some animal action might be embedded within a compensatory structure such that the behavior is guided from start to finish (e.g., a leopard killing its prey), other animal action might be more of the point-and-shoot variety (e.g., a frog catching a fly with its tongue).[11] Or in the actual world, there may be no such thing as irreducible goal-directedness, irreducible agent-causation, or two-way powers and so no actions explained by such phenomena; but presumably such phenomena *could* exist, and in possible worlds where such phenomena existed, presumably they would pick out actions rather than mere behaviors.[12]

In light of these reflections, the potential differentia coming from many existing perspectives on agency—e.g., causation by belief-desire pairs, causal sustenance by intentions, being under the guidance of the agent, being a

(Sartorio 2010). For example, I might not go to class today because I am feeling sick; in omitting to go to class, I am performing an action; yet it looks like this action is not constituted by a movement. While I do not have space to develop this idea in detail here, it is important to remember that according to my terminology, a movement can involve only mental changes. Thus, I conjecture that we see omissions that are actions as (partially) constituted by mental movements, e.g., decisions or resolutions to not do something (this proposal is in the spirit of Shepherd 2014*a*). Resistings, sustainings, and allowings that involve no physical changes could be dealt with similarly, I conjecture.

[9]For a similar thought, see O'Brien's (2017) argument from multiple sources. See also Milgram 2010 and Milgram 2022; in the latter, we get an evolutionary explanation for why there are different forms of agency too.

[10]This example comes from Lewinsohn 2024.

[11]For examples of action of the point-and-shoot variety amongst humans, see Kelley 2024. See also Churchland 1986, pp. 430-431 and Rosenblueth et al.'s (1943) discussion of non-feedback purposeful behavior.

[12]To be clear, it may be that *intentional* actions *in our world* are marked out by one particular sort of aim or one particular sort of explanatory structure. Plausibly, understanding the distinguishing feature of intentional actions in our world has been the goal of many existing theories of action. But this is not my goal. I am interested in understanding the mark of (reflexive) action, whether or not the (reflexive) action is intentional and whether or not it is actual.

manifestation of practical knowledge, being irreducibly goal-directed, being irreducibly agent-caused, and being a manifestation of a two-way power—are predictably at the wrong level of generality to solve the problem of action as I understand it. They might be at the right level of generality to be marks of (actual) *intentional* action, but as marks of (actual and merely possible) action, they are all too narrow. None is a feature that is present in all instantiations of action in all possible worlds.

These basic intuitions about the generality of action lead to an idea and a conjecture. The idea: what we should instead be looking for in trying to solve the problem of action is what *unites* all of these agentive phenomena in delineating forms of action; and the conjecture: what unites these phenomena is that each is a form of *reflexive control*, that is, control an individual has over its own movement. Thus, assuming that what delineates a form of action should be a form of what delineates action, we conclude that what delineates action is reflexive control as such: action is reflexively controlled movement.[13] Forms of reflexive control then correspond to forms of action.

To summarize, here's an *argument from the generality of action* for the conjecture that action is delineated by reflexive control.

P1. Each kind of action involves a distinctive kind of reflexive control.

P2. Each kind of reflexive control delineates some distinctive kind of action.

P3. So no kind of reflexive control could delineate action.

P4. What delineates each kind of action is a form of what delineates action.

C. So action is delineated by reflexive control.

To precisify this argument and the resulting control theory of action, we need a more precise account of what it is to control one's own movement.

3.3. Reflexive Control

We begin with the thought that reflexive control over movement is a special case of *movement control*, where movement control consists of an individual x controlling the movement of an individual y; *reflexive (movement) control* then consists of movement control where x is identical to y.[14] So we will make progress understanding reflexive control by understanding what it is for two individuals x and y to stand in the relation of x controlling

[13] Control has played a central role in recent discussions of the nature of intentional action (as in Shepherd 2014*b*, Shepherd 2021, Wu 2016, Smith 2021, Pavese 2021, Beddor and Pavese 2021, and Valaris 2022).

[14] I will use 'control' to refer to movement control.

a movement of y. Another motivation for setting up our inquiry in this more general fashion is that controlling one's own movement is intimately connected to controlling another individual's movement (e.g., I control my turning left at an intersection by controlling my steering wheel rotating left), and so it will be useful to have a theory of both on the table.

To offer an account of movement control, I start by highlighting three formal features that a theory of movement control should capture. First, a movement is controlled relative to a description, feature, or aspect of the movement (Wu 2016; Wu 2023). For example, I might control my arm movement under its description 'raising my arm' but not under its description 'knocking over a cup'. Thus, we will speak of x controlling a movement m of y *relative to description* φ; I call a movement that falls under description φ a φ*ing*. Second, control is gradable. For example, I can come to control my backflip more and more with practice, or I might control my dog's barking to a greater degree with training. This feature of control also entails that when speaking about control as a binary property, there is an implicit reference to a threshold of sufficient control. Third, being controlled is a *determinable* property of movements. Roughly, a property P stands in the determination relation to a property P' if P increases the specificity of P'; to be P is to be P' but in a specific way; a determinable is then a property relative to which another property—a determinate—stands in the determination relation.[15] And the determination dimensions of a determinable are the dimensions along which a determinable can be made more determinate (Funkhouser 2006).[16] In too many instances to cite, control is treated, either explicitly or implicitly, as a single kind of thing; and often, theories of control mistakenly identify control (per se) with a *kind* of control.[17] But I suggest that we start to understand control once we explicitly recognize that being controlled is a determinable property that takes more determinate forms along particular determination dimensions.

I now offer a more precise characterization of movement control which illuminates these three formal features. To preview, movement control involves the satisfaction of a trio of conditions: an *aim condition* that captures the sense in which control involves a success, a *modal condition* that ensures the success is non-accidental, and an *explanatory condition* that requires that the non-accidental success is attributable to the controller.[18] Let me take each of these conditions in turn.

[15] See Chisholm 1985 and Ford 2011, pp. 83-88 for helpful discussions of the determination relation in the context of philosophy of action. See Prior 1949, Funkhouser 2006, and Wilson 2021 for more general treatments.

[16] The quintessential example of a determinable is being colored, with being red, being blue, being yellow, and so on as determinates; to be red is to be colored but in a specific way. The determination dimensions of color are thought to be hue, brightness, and saturation. Having taste is another example of a determinable, with modalities of taste such as being sweet and being sour as determinates.

[17] See Pavese 2021 for criticism of multiple theories of what she calls 'agentive control' for making the latter mistake. While I will not argue for this here, I think Pavese makes the same mistake as those she criticizes: her epistemic theory of agentive control in terms of knowledge is, at most, a theory of a *kind* of reflexive control.

[18] These conditions should not feel unfamiliar. What is most novel about my theory of control is that it puts together the three conditions in general enough forms to account for the fact that control is a determinable.

3.3.1. Aim condition

The first necessary condition on x controlling y's movement relative to description φ is a condition on the internal state of the controller: during y's φing, x in some sense aims at y φing.[19] One might *influence* or *cause* the movements of one's steering wheel when one aimlessly hits it with one's arm while waving at a friend, but in this case one does not *control* the movement of the steering wheel. One difference between control, on the one hand, and influence or causation, on the other, is the involvement of an aim. It is also through aiming that control incorporates success: when x controls y's movement, x sets a standard on y's movements such that when the standard is met, x in some at least minimal sense succeeds.[20]

What is it to aim at y φing in the sense relevant here? The first thing to say is that aims come in different forms. An aim might consist in an intention (Bratman 1987), a motivating reason (Davidson 1980*a*), a piece of practical knowledge (Anscombe 1957), a value (Watson 1975), a normative judgment of goodness (Tenenbaum 2007), or a structure of interrelated attitudes rather than a single attitude (Frankfurt 1971). Second, the description under which one aims at a movement might be more or less detailed. One might aim simply at y φing or one might aim at y φing by way of a specific means. In the latter case, one's aim is met and one control's y's φing to the degree that both the target movement and the target means are realized.[21] Third, one need not aim at y's movement under a bodily or mental movement description. One might aim at y starting a war, graduating from a doctoral program, or building a birdhouse. Also, one can have either a *de re* aim of an ongoing movement that it be a φing or a *de dicto* aim that y perform a φing (Wilson 1989). Fourth, I do not assume that aiming requires a sophisticated capacity for qualitatively rich, linguistically infused mental states; nor do I assume that all aims are conscious. I mean to make room for many animals having the capacity for control and thus the capacity to aim (Tomasello 2022). Moreover, I want to capture that you in some sense aim at the precise movement types you token while performing a skilled golf swing, even if these aims are subconscious or automatically generated. So, for example, motor representations deployed by conscious intentions should count as aims in the relevant sense.

Here is a proposal for a definition of aiming which meets these four desiderata: to aim at y φing is to represent y φing in a positive or to-be-realized manner.[22] I take it that this account of an aim also fits relatively

[19] As Dennett (1984) says, 'for something to be a controller its states must include desires—or something "like" desires—about the states of something (else)' (p. 58).

[20] The aim condition also helps explain why control is description or feature relative: x might aim at certain features of y's movement and not others.

[21] So the aim condition helps explain why control is gradable. The modal and explanatory conditions will also influence the degree of control x has over y's movement.

[22] For more on map-like and nonconceptual representations—and thus the possibility of aiming in less linguistically and psychologically sophisticated animals—see Camp 2007 and Pacherie 2011, respectively.

well with our pre-theoretical understanding of an aim, but insofar as it does not, one is free to treat 'aim' here as a technical term.

Importantly, x controlling y's movement relative to description φ requires that it be x rather than someone or something else that represents y φing; thus, while an aim might be subconscious and automatically generated, it must in some sense still *belong* to x rather than a subsystem of x. The aim's satisfaction must be a success *to the agent*. This leaves room for subconscious aims, because a subconscious pro-representation—such as a motor representation—deployed in service of one of x's conscious aims is plausibly an aim that belongs to x.[23] What more exactly does it take for a representation to belong to x? Answering this question is a difficult task, one which I do not intend to settle here, though I will say a bit more in Section 4. It is difficult because it interacts with challenging questions about personal identity and the individuation of entities.

Leaving the notion of 'belonging' as a placeholder for further theoretical work then, here is the more precise formulation of the aim condition:

> (**aim condition**) during y's φing, x aims at y φing, where this requires that x represents y φing and the representation
>
> (a) belongs to x and
>
> (b) represents y φing in a positive or to-be-realized manner.

3.3.2. Modal Condition

When x controls a movement m of y relative to description φ, x not only aims at y φing, but this aim is connected to y's φing in some distinctive way. If I form an intention to move the steering wheel to the left and immediately afterwards a strong wind turns the wheel to the left, my representation is not connected to the movement of the wheel in the way required for me to have controlled the movement. The key question is how exactly my aim must be connected to the movement in order for me to control it. Again, we should expect this answer to be sufficiently general such that it can be precisified in multiple ways.

The first thing to say is that the connection should entail that the meeting of the aim—the success involved in movement control—is not an accident. Put another way, accident or luck is antithetical to control. But there are different ways that success might be non-accidental. The first way is captured by Frankfurt's (1978) guidance theory of action: the occurrence of a φing is non-accidental if its occurrence is not dependent on the precise features of the situation but instead would occur even in a variety of slightly different circumstances in

[23] For instance, when a dancer consciously aims to improvise a dance, the motor representations triggered by the conscious aim to improvise would count as the dancer's aims.

which one's aim remains the same (e.g., circumstances with barriers to success in place).[24] Call a movement satisfying such counterfactuals *circumstantially robust*. The more circumstantial robustness, the less the success was an accident and the more control had over the φing.

A second way that meeting x's aim might be non-accidental is if the features of y's behavior are sensitive to the content of x's aim. In light of this sensitivity, the fact that there is a match between y's behavior and x's aim is, in a sense, no accident. This sense of non-accidentality is captured by a certain form of counterfactual highlighted by Bishop (1989): had x aimed that y φ' instead, y would have φ'ed. Call a movement satisfying such counterfactuals *content sensitive*. The more content sensitivity, the less the success was an accident and the more control had over the φing.

To get a clearer picture of these two forms of non-accidentality and the kinds of control they capture, imagine two basketball players. The first can make a free throw in lots of different circumstances; the air could be colder, she could be a bit more tired, and the court could be indoors or outdoors. So her free throw is circumstantially robust. The second can make many kinds of free throws; she could swish, hit it off the backboard, bounce it off the front rim, and bounce it off the top of the hoop, but only in very precise circumstances. So her free throw is content sensitive (but not circumstantially robust). Both basketball players meet their aim to make a free throw, both *non-accidentally* meet their aim, and both significantly control their free throw—albeit in different ways.[25]

A condition that naturally generalizes both circumstantial robustness and content sensitivity requires that the φing have sufficient *modal support*, where modal support comes in the form of true *anti-luck* counterfactuals. An anti-counterfactual is a counterfactual of the form: *had circumstances been different in such and such sufficiently small ways and/or the content of x's aim been different in such and such sufficiently small ways, y's behavior would match x's aim.* We are thus led to the following modal condition on control in these terms:

(**modal condition**) the φing is sufficiently modally supported.[26]

[24] See also Aguilar 2012, Shepherd 2021, and Valaris 2022.

[25] One might argue that in the second example, there is still circumstantial robustness but very little of it. Perhaps, but if that is *all* that can be said, then this would seem to imply that control is significantly lacking in this case. This would be misleading, however, since the extreme flexibility that the player has in the very particular circumstances she is actually in is an indicator of significant control, just of a different kind than the kind one has when one's success is largely circumstantially robust. So we should allow that different modal profiles are associated with different forms of control.

[26] I state the modal condition in terms of counterfactuals but take no stance on the many theoretical issues surrounding counterfactuals. In particular, I take no stance on whether we should understand the truth conditions for anti-luck counterfactuals in terms of, e.g., a Lewisian semantics or a probabilistic semantics using objective chances (Hájek ms, Leitgeb 2012). Moreover, I am not particularly tied to appealing to counterfactuals as opposed to the corresponding conditional chances in an account of non-accidentality. There is historical precedent for appealing to counterfactuals in a theory of action (especially a theory of free action), but it would be worth thinking through whether conditional chances might be better suited. For a theory of control in terms of success-rates, so something a bit closer to objective chances, see Shepherd 2021. Thanks to Al Hajek for helpful discussion.

3.3.3. Explanatory Condition

The aim and modal conditions discussed thus far are not enough to ensure that there is control. To see this, consider digestion. Digestion is constituted by bodily movements that are embedded within a homeostatic system and thus are supported by a number of true anti-luck counterfactuals. Now imagine that as you feel your stomach gurgling, you form an aim to digest. In this case, you do not control the digestive movements under any description, despite meeting the aim and modal conditions. The reason for this is that your aim is not part of the explanation of the modally supported digestion. Because of this, when you aim at digesting and non-accidentally digest, the non-accidental success of your digesting is not *attributable* to you. Thus, the first part of the explanatory condition on a movement of y being controlled by x relative to description φ is that x's aim that y φ be a part of the explanation for the movement being a modally supported φing (Wu 2023, p. 29).

But adding just this explanatory condition does not suffice either.[27] Imagine that you form a present-directed intention to kick your leg. A doctor doing an examination on you is, unbeknownst to you, a mind reader and knows about your present-directed intention to kick your leg. Wanting to fulfill your intention, he immediately taps your knee with a little hammer causing your knee to go up. Further, the doctor is skilled such that he would manage to tap your knee despite many changes in circumstances and potential obstacles. Who controls the circumstantially robust kick of yours in this case? You or the doctor? Presumably the doctor does; yet both you and the doctor aim at the kick and both of these aims are part of the explanation for the movement being a modally supported kick.

Again, we need to look at the nature of the explanation of the modally supported kick. One might think that the reason why you do not control your kick is simply that this explanation is constituted by too many facts about the doctor. And indeed, one way for x to meet the second part of the explanatory condition on controlling y's φing is for the explanation of the modally supported φing to be *internal to* x, that is, sufficiently constituted by facts about x. Which facts about x? I require that these facts be facts about x's body and/or mind, and that's it.[28] One might worry that this permissive of an explanatory condition leaves room for deviant causal chains to suffice for control when they operate entirely within the agent's body and mind. For instance, when Davidson's (1980*b*) climber intends to drop his climbing partner, and the intention upsets him so much that his hand shakes and he drops his partner, the explanation of the (minimally modally supported) dropping includes his intention to drop his partner and is entirely constituted by facts about the climber. Thus, on the

[27] Peacocke's (1979) neurophysiologist case is relevant here.

[28] The significance of facts about x's mind and/or body is that I am assuming that an individual is exhausted by its mind and body (recall Footnote 5). If x is made up of parts beyond its mind and body, then facts about those parts also suffice to meet the internal component of the explanatory condition.

proposed theory of control, the climber controls his dropping his partner.

There is a more concessive response and a less concessive response to this objection. The more concessive response is to point out that something like the problem of deviant causal chains is a central problem for all sorts of philosophical theories that appeal to explanatory notions. If it shows up in my theory of control, it shows up in the form of needing to specify more precisely the sort of facts about x's body and/or mind that must explain a non-accidental φing for x to control the φing. The less concessive response is that some causal chains usually described as 'deviant' suffice for control on my theory, but this is the right result given that I am interested in control as such. Davidson's climber does *in some sense* and *to some small degree* control the dropping of his partner. It's just not the kind of control required for intentional action, and there isn't much of it.[29] Indeed, the drop constitutes a non-accidental success by the climber's own lights. After all, he did intend to drop his partner, and he would have got nervous and dropped him in a range of nearby circumstances in which intended to drop him, so dropping him wasn't an accident or a fluke.[30] Moreover, the non-accidental success is in some sense attributable to the climber, in particular, to his shaky hands. Thus, it would be odd to describe the dropping of his partner as entirely and in all ways out of the climber's control.[31] In my opinion, this deflationary response to the problem of causal deviance is the right one. It is principled and intuitive. But if one is not convinced, one can accept the more concessive response and replace 'facts about x's body and/or mind' with 'the right sort of facts about x's body and/or mind' in the explanatory condition below.

Let us return to the doctor example. While sometimes control is manifested internally, this need not be the case. So the problem in the doctor example is not *just* that the explanation of your kick is external to you. Indeed, agents can control movements by way of reliable mechanisms in their surroundings, including machines and other agents (Dennett 1984, pp. 59-62); but in such a case where the explanation for the movement and its modal support is largely external to the agent, for control, the agent needs to *make use* of these external mechanisms as means to achieving their aim, not just be in the right place at the right time. I assume that one makes use of an external mechanism as a means when i) one aims at that external mechanism being (part of) the means by which one realizes a relatively more intrinsic aim and ii) this instrumental aim plays an explanatory role in the realization of one's relatively more intrinsic aim by way of the external mechanism.

So in short, the reason you do not control your kick in the doctor example is that you do not make use of

[29] Because there isn't much of whatever kind of control the climber has, the climber dropping his partner may not be an action on the control theory. Control is gradable, and so it depends on where the threshold of sufficient control is set.

[30] There is also a small degree of content sensitivity. For instance, had the climber intended to not drop his partner, he wouldn't have.

[31] Since the climber didn't intend to drop his partner at the very moment he did, the climber did not control the time at which he dropped his partner (or, put more precisely, he didn't control the dropping relative to a description which specifies the time at which the dropping occurred). But one can control one's φing even if one doesn't control when one φs. For instance, I might control my winning a tennis match, but I don't control the time at which the win occurs.

the doctor as a means to kicking. Imagine instead that you had known about the doctor's ability to mind read and interest in bringing about what you intend in a circumstantially robust manner. Had you then intended to kick by way of the doctor's knee tap and had this intention played an explanatory role in your kicking by way of the doctor's knee tap, then *both* you and the doctor would have controlled your kick. (Note though that the control over the kick attributed to you and to the doctor would be of different kinds and degrees.)[32]

We thus arrive at the full explanatory condition on x controlling y's φing.

> (**explanatory condition**) the facts that constitute the explanation for the movement of y being a modally supported φing (i) include the fact that x aims at y φing and (ii) are either (a) *internal* to x, i.e., composed to a sufficient degree of facts about x's body and/or mind or (b) *external* to x, i.e., not internal to x, but x makes sufficient use of these external facts as a means by which to φ.

Call an explanation of a modally supported φing *x-centered* insofar as it satisfies (i) and either (ii)(a) or (ii)(b) relative to x.

3.3.4. The Full Account

Here is the final proposal for necessary and sufficient conditions for movement control.

> (**movement control**) x controls a movement m of y relative to description φ just in case i) x aims at y φing during the φing, ii) m is a modally supported φing, and iii) the explanation for the fact that m is a modally supported φing is x-centered.

These three conditions correspond to the determination dimensions of movement control, which we will label aim, modal support, and explanation, respectively. To give a precisification of aim, one specifies a kind of aim; to give a precisification of modal support, one specifies the kind of anti-luck counterfactuals that are to be satisfied; to give a precisification of explanation, one specifies a kind of centered explanation for the movement and its modal support.

Finally, using the relationship between reflexive control and movement control, we get necessary and sufficient conditions for reflexive control.

> (**reflexive control relative to φ**) x reflexively controls a movement m of theirs relative to description φ just in case i) x aims at oneself φing during the φing, ii) m is a modally supported φing, and iii) the explanation for the fact that m is a modally supported φing is x-centered.[33]

[32] Moreover, the descriptions under which you and the doctor controlled the kick would likely be different. While you control the kick under the more general description 'kick', it might be only the doctor that controls the kick under the more specific description 'very forceful kick'.

[33] Notice that reflexive control requires the content of one's aim to refer to oneself as such. This minimal self-conception in the

3.4. Action as a Determinable

Let us return to the problem of action. With the argument from the generality of action precisified, we tentatively conclude that action is reflexively controlled movement (I say tentatively because I must still respond to three objections in Section 5). Reflexive control is a determinable whose determination dimensions correspond to the aim, modal, and explanatory conditions. It follows that action is a determinable with the same determination dimensions. More precisely:

> (**action relative to** φ) a movement m of x is an action of x relative to description φ just in case it is reflexively controlled relative to description φ.
>
> (**action**) a movement m of x is an action of x just in case it is reflexively controlled relative to some description.[34]

These two definitions answer two different questions underlying the problem of action. The first answers the question 'when does one φ as an action?', a question that involves a salient description; and the second answers the question 'when is a movement an action?', a question that does not involve a salient description.[35]

The control theory of action also offers a *schema* for theories of particular forms of action. To give an account of a particular form of action—say, intentional action—from the perspective of the control theory, it is necessary and sufficient to give an account of:

- The kind of aim which is distinctive of intentional action;
- The kind of anti-luck counterfactuals which are true of intentional action;
- The kind of centered explanation which an intentional action is subject to;
- The threshold of sufficient control determining the extensional boundary of intentional action.[36]

So the fact that action is a determinable on the control theory illuminates the point about the generality of action discussed in Section 3.2; and in isolating the determination dimensions of action, the control theory allows

content of the representation which constitutes an aim can be highly implicit and need not involve a sophisticated conception of self. For example, if one aims *to* φ, the use of an infinitive involves the required implicit reference to self. Moreover, the aim need not be embedded within a qualitatively rich, linguistically infused mental life. Still, in some at least minimal sense, x does not only aim at x φing but does so with the understanding that the behavior it aims at is its own.

[34] I have defined action in terms of reflexive control. But if action were defined in terms of movement control over a wider class of movements than the agent's own, one would have a version of the control theory consistent with the extended action thesis mentioned in Section 3.1. It is in this way that the core components of the control theory are neutral with respect to the extended action thesis.

[35] Thus, much like Davidson's causal theory of action, the control theory of action captures that action has two guises: a description dependent guise and a description independent guise.

[36] Since the control theory of action entails a schematic theory of intentional action, if a theory of intentional action doesn't fit the proposed schema, it is in inconsistent with the control theory. But as the table below shows, many prominent theories of intentional action can be made consistent with this schema.

us to think systematically about this generality. In particular, it provides resources to systematically compare and contrast different kinds of action.

At this point, we are in a position to see how the control theory of action offers a unification of many existing theories of action. The idea is that we can see many existing theories of action as isolating forms of reflexive control and thus determinates of action. In particular, we can see them as specifying particular ways of meeting the aim, modal, and explanatory conditions. In the box below are four examples of how we might see existing theories of action as isolating determinates of action.

Davidson's causal theory of action

- aim = belief + desire pair
- modal support = content sensitivity
- explanation = causation by belief + desire pair

Wilson's teleological theory of action

- aim = de re intention
- modal support = content sensitivity
- explanation = irreducible goal-directedness

Bratman's planning theory of action

- aim = intention as plan state
- modal support = circumstantial robustness
- explanation = causal sustenance by intention

Frankfurt's guidance theory of action

- aim = ?
- modal support = circumstantial robustness
- explanation = compensatory causal mechanism

To be sure, the theories of Davidson, Wilson, and Bratman were developed to be in competition with one another as theories of intentional action. On the control theory of action, intentional action is identified by a distinctive sort of reflexive control—what we might call *intentional control*. Thus, from the perspective of the control theory of action, we can view the three theories as competing theories of intentional control. Even if only one of the three theories succeeds as a theory of intentional control, the table shows that they all succeed in isolating *some* form of reflexive control. Thus, each succeeds in identifying some determinate of action (whether or not that determinate is intentional action). In this way, from the perspective of the control theory of action, there is a new way of thinking about competing theories of intentional action: as each identifying some potentially interesting determinate of action.[37]

One might think that Frankfurt's guidance theory of action, on the other hand, is a competing account of reflexive action. Indeed, when he sets up the problem of action, he focuses on a phenomenon connected to

[37]The same point applies for theories of autonomous action, free action, and other kinds of reflexive action.

the category of action that is narrower than the most minimal forms of agency but broader than distinctively human forms of agency such as intentional and autonomous agency. However, as a theory of reflexive action, the guidance theory fails in two ways. First, without an aim conditions, it is hard to see how his theory captures that all reflexive actiosn constitute a success by the agent's own lights. Second, it assumes that non-accidentality takes the form of circumstantial robustness, but as we saw, non-accidentality can also take the form of content sensitivity.[38]

4. Examples

So that is the control theory of action. Its central commitment is that action is reflexively controlled movement, where being reflexively controlled requires meeting an aim condition, a modal condition, and an explanatory condition. Moreover, being reflexively controlled is a determinable property of movement and so being an action is a determinable property of movement as well. This feature of action—that it is a determinable—illuminates the diversity within the category of action and allows for a unification of many existing perspectives on agency.

To give a bit more concreteness to the control theory of action, in this section I consider a number of example behaviors and discuss what the control theory has to say about them. While much more needs to be said about certain cases, I hope the reader will nonetheless benefit from the brief remarks.

1. **Failed action** (e.g., missing a free throw): the event of missing a free throw is an action under the description, e.g., 'attempting a free throw' but not under the description 'making a free throw' nor under the description 'missing a free throw'. There is no event at all that falls under the description 'making a free throw', and one does not aim to miss the free throw so the aim condition is not met with respect to that description. The event of missing a free throw is then an action (full stop) because it is reflexively controlled under some description.[39]

2. **Unintentional action** (e.g., unintentionally insulting someone): the event of insulting someone is an action under the description, e.g., 'saying XYZ' but not under the description 'insulting someone' since one fails the aim condition with respect to the latter description. So the event of insulting someone is an action because it is reflexively controlled under some description.

[38] Regarding the first point, I conjecture that this gap in Frankfurt's theory is symptomatic of never fully developing what it takes for guidance to be *the agent's* guidance. A plausible account of what makes guidance the agent's guidance is that the guidance is directed by an aim of the agent.

[39] Recall that I started with the idea of actions that involve a success by the agent's own lights. The theory of this target joint now clarifies that all such actions constitute *some* success but may constitute a failure *in addition*. What about behavior that radically fails to meet any of one's aims? Well, usually when you aim to φ and then go on to attempt to φ, you also aim to attempt to φ; and so these cases of radical failure are going to be strange. Nonetheless, I do not want to rule them out as impossible. For my purposes here though, all I need to be plausible is that such behaviors are not *reflexive* actions, only actions in a more minimal sense like the sense that is interchangable with *something doing something*.

3. **Fidgeting** (e.g., drumming one's fingers): whether forms of fidgeting are actions depends on the exact psychology behind the behaviors—whether they are explained by aims, albeit subconscious ones—and to what extent we think these subconscious aims belong to the agent. Answering these questions requires further empirical and philosophical work.

4. **Reflexes** (e.g., jumping at a scary movie): since reflexes, as I understand them, involve responding directly to stimuli without the involvement of an aim, they are not actions.

5. **Chancy action** (e.g., hitting bullseye): hitting a bullseye while aiming to do so is not circumstantially robust assuming the thrower is unskilled; still, depending on the threshold of sufficient control, hitting the bullseye will satisfy the modal condition because hitting the bullseye is minimally content sensitive; for example, had one aimed to instead throw the dart at the ground, one would have done so. With that said, such true counterfactuals require a relatively large change in the content of the aim—had one aimed to hit the bullseye slightly later, for example, one would not have. Thus, hitting the bullseye is an action under that very description in contexts where the threshold of sufficient control is low.

6. **Weak-willed action** (e.g., the unwilling addict buying a drug): on the control theory, weak-willed action is behavior that is controlled in one sense but not another; when the drug addict buys the drug because of their desire for the drug, in the usual case, they meet the three conditions needed for control over their buying the drug; with that said, presumably they do not endorse their desire—they feel alienated from their desire—and so they fail to aim at buying the drug in a more substantive sense; thus, they fail to control their buying the drug in a more substantive sense connected to the more substantive sense of aiming; still, buying the drug is an action (under that very description) because they have some form of control over it.

7. **Fine motor skills** (e.g., a precise movement of one's finger while playing Moonlight Sonata): assuming the finger movement is brought about by a motor representation deployed in service of an intention to play Moonlight Sonata and whose content specifies the precise details of the movement, the movement plausibly meets the aim condition (under its most detailed description specifying its speed, direction, and so on). We can also set the case up so that the modal and explanatory conditions are met. In such a case, the finger movement will count as an action under its most detailed description.

8. **Freudian action** (e.g., subconsciously sabotaging a friendship): sometimes our behavior is influenced by motives that we do not have conscious access to; yet conscious access to an aim isn't required for the

aim to be one's own—the subconscious aim may, for example, be deeply integrated with the person; so if one sabotages a friendship, motivated by some such subconscious aim (and the modal and explanatory conditions hold), the sabotaging will count as an action under that very description.

9. **Foreseen side-effects** (e.g., intending to bomb an ammunition plant, while anticipating but not intending civilian causalities): foreseeing but not intending a side-effect of what one intends counts as an aim in my sense (there is a clear sense in which the side-effect is seen by the agent as something to be brought about); thus, bringing about civilian casualties as a foreseen but unintended side-effect of bombing the ammunition plant is an action under that very description.

10. **Goal-directed activity of non-sentient systems** (e.g., HVAC system): it is unclear whether such behaviors are actions, because it is unclear whether non-sentient individuals like HVAC systems can aim in the sense of having pro-representations or if, rather, we merely *speak* as if they have aims (Tomasello 2022, p. 18). Moreover, even if HVAC systems do aim in the sense of having pro-representations, it is unclear whether these aims belong, in the requisite sense, to the non-sentient system or rather to the sentient individual who built the system.

11. **Biological functions** (e.g., digestion): whether certain cases of biological functioning like digestion count as actions depends on the content of one's aim and the explanatory connection between that aim and the biological functioning. If you aim to digest your food after consuming it, then the aim will not be explanatorily relevant to your in fact digesting and so the explanatory condition will not be met. Say, on the other hand, you consume food while aiming to digest your food and this aim explains your eating and thus digesting. Then your digesting may be an action.[40]

This list also illustrates one way that intentional action is narrower than reflexive action: the former but not the latter requires a degree of conscious access to what one is doing. One cannot intentionally φ while having no idea that one is φing.[41] Thus, two places that intentional action and reflexive action come apart is with respect to fine motor skills and Freudian action. Given a piano player might have no idea what they are

[40] For a defense of the idea that a behavior one merely causes by performing another action could be an action, see Kelley 2024. Still, one might find it an unwelcome result that digestion could constitute an action—perhaps even *usually* constitutes an action insofar as we usually eat with the aim of digesting. The first thing to say is that digestion is an action of a different kind than, e.g., your raising your arm. And this can be captured by the difference in the way the explanatory condition is met. Digestion can be an action more in the way that losing weight can be an action. Second, note that there is one sense in which your digestion is out of your control and thus not a matter of agency: you do not control, e.g., the token digestive contractions of your stomach. This is because even if you aim at these contractions and your aim to so contract explains your digesting, your aim does not explain the particular contractions involved in the digestion. This is analogous to how the lightening strike might explain the burning down of the house without explaining the particular sub-components of the burning down of the house. So in this way, all you control and all that is an action are your digestive movements under very general descriptions.

[41] See, e.g., Alvarez 2009. This is not to say that one must *know* that one is φing.

doing with their finger at a particular moment in playing Moonlight Sonata, it would be a stretch to call the movement intentional (under the precise description that is consciously inaccessible).[42] Similarly, a Freudian action, while reflexive, is plausibly not intentional—again, in such cases, we may assume that one has no idea that one is engaged in the Freudian action, one would disavow that one aimed to perform the action, and so on.[43] Another difference between intentional and reflexive action is with respect to bringing about foreseen but unintended side-effects. There is widespread debate about whether such doings are intentional, but many think they are not; so such doings are at best borderline cases of intentional action. However, such doings are clearly reflexive actions.

5. Objections

In this section, I will address three objections to the control theory of action. All three objections are precisifications of a more general worry that once it becomes clearer what exactly control is, it becomes less clear that there is as tight of a connection between control and action as the control theory suggests. The three worries target the three conditions of control, respectively.

The first worry takes us back to earlier discussion: one might think that even if control requires a representation the belongs to the controller, action does not require a representation that belongs to the agent. As discussed at length already, I have conceded for the purposes here that there could be minimal forms of action which do not involve an aim, e.g., the paramecium's reproduction or the sun's melting the chocolate. I restricted myself to defending the control theory as a theory of reflexive action, where this phenomenon is (prima facie) narrower and is demarcated by the idea of a success by the lights of the agent. And the aim condition is exactly the kind of condition needed to ensure that a standard is imputed by and met by the individual when they act in this sense. Moreover, it's not clear in what way other than via a representation belonging to the agent a standard of success could be placed on the agent's movement by the agent.[44]

Second, one might worry that the modal condition on action is too strong such that even if control is a modal notion, action is not. In a sense, this thought goes back to Frankfurt's objection to the Principle of Alternative Possibilities (Frankfurt 1969). He claims that even if your doing is to a large extent not content sensitive such

[42] See Brozzo 2017 for an argument that some motor representations can be consciously accessible. Still, in cases where the guiding motor representation isn't consciously accessible, there opens a gap between intentionality and reflexivity.

[43] Marcus (2019) argues that actions involving self-deception are intentional. Even if this is right, there are Freudian actions that do not involve self-deception. For example, Velleman (1992) offers a case where one's motive crystallizes over time without one's noticing.

[44] As David Heering pressed me on, perhaps success by the lights of the agent could alternatively be understood in terms of the exercise of an ability. This proposal deserves more attention than I can give it here, but my preliminary response is that more needs to be said to motivate why the exercise of an ability in itself—independent of whether the exercise is in service of an aim of the agent—is a success by the lights of the agent.

that you could not have done otherwise even if you wanted to, that doing might still be an action for which you are morally responsible. Of course, this point is compatible with the control theory of action since content sensitivity is not required for action. However, there is a related worry that is a challenge to the control theory. A first pass at the worry is a purported counterexample: imagine that you raise your arm because you aim to do so; however, there is a counterfactual intervener that remains inactive in the actual world but would have thwarted your attempt had the circumstances or your aim been even slightly different. In this case, it looks like the modal condition fails. Yet it looks like you raised your arm as an action.

It is not right to say that the modal condition fails in this example, however, since we must ask what would happen in circumstances without the counterfactual intervener around. Part of what is implied by the fact that your aim explains your raising your arm is, I take it, the truth of a number of such anti-luck counterfactuals. Thus, this is really an example of *deficient* control, since the true anti-luck counterfactuals that support the movement involve more distant possibilities. Yet our intuitions about the case suggest that the control is enough for the arm raise to be an action.

One might press the objection further by removing the counterfactual intervener and imagining a case where for virtually *any* change in circumstance and virtually *any* change in content, your action would fail to match the content of your aim. The conditions of the actual world are essentially the unique conditions needed for you to succeed at your aim. Despite the success being in all senses lucky and thus uncontrolled, shouldn't your arm raise nonetheless be an action? The deeper thought underlying the objection at issue thus seems to be the thought that what matters for action is not what *would* have been the case but what *actually* is the case.

However, this thought problematically presupposes a strict division where there is not one: what actually is the case is inextricably connected to what would have been the case. Causation, for example, is often thought to be intimately connected to the truth of counterfactuals, in particular, content sensitivity (Lewis 1973; Lewis 2000). Guidance of a movement by an agent is often thought to entail certain modal facts (Frankfurt 1978; Fischer and Ravizza 1998). Irreducible goal-directedness is thought to support a number of true anti-luck counterfactuals (Wilson 1989, ch. 7). Put another way, one thing seems clear about (reflexive) action: for a φing to be an action, the agent's aim to φ needs to at least partly explain their succeeding to φ. But most of the proposals for what this explanatory relation more precisely consists in have modal implications, often of the kind needed to satisfy modal support; perhaps this is even because explanation itself is modal in character. For these reasons, we should: i) be skeptical that the purported counterexample at issue is an action, for given that the arm raise has essentially no modal support, we should be skeptical that there an explanatory relation between the arm raise and the agent's aim, ii) be amenable to the thought that action is a modal notion, and iii)

see the strong connection between action and control on the control theory as capturing this feature of action.

The third and final objection I will consider is about allowing action via external forms of control. While it might be admitted that one can control one's movements via the actions of another, one might insist that one's *action* cannot be performed by way of another's action. You cannot, for example, kick your leg as an action by way of your doctor intentionally tapping below your knee; and this is so even if you could cause your kick or control your kick by making use of your doctor's intentional tapping.

My response here is a conjecture: we will have a more joint-carving and less gerrymandered picture of action if we make room for the exploitation of another's agency in action.[45] Consider the following three ways of kicking:

(1) kicking by way of the internal mechanisms of one's own body (as one usually does);

(2) kicking by way of some external mechanism;

(3) kicking by way of another agent's action.[46]

It is difficult to see why (1) is an action but (2) isn't, for we make use of external mechanisms to act all the time. Consider, for instance, mailing a letter cross-country using USPS, translating a Spanish text into English using a Spanish-English dictionary, or lifting a building off of the ground using a crane. Perhaps the worry is instead that we cannot make use of external mechanisms to perform actions relative to bodily movement descriptions. But given that we can perform actions relative to other descriptions by making use of external mechanisms, what stops this from being the case when the description makes use of bodily movement terms?

One might answer this question by suggesting that bodily movement terms like 'kick' can only be instantiated as actions when performed basically, i.e., without performing another action as a means. So by the meaning of the word 'kick', a kick cannot be performed as a nonbasic action; but making use of an external mechanism to kick requires kicking as a nonbasic action.[47]

First, making use of an external mechanism to kick does *not* require kicking as a nonbasic action. When you φ as a nonbasic action, you φ by performing some other action ψ as a means. But the operation of the external mechanism need not require another action of yours. For example, the recent development of neural interfaces allows paralyzed persons to control the movements of their (prosthetic) limbs by way of a connection between the neural interface implanted in their brain, a nearby computer, and their (prosthetic) limb (Jabr 2022). In

[45] This claim is consistent with the falsity of the extended action thesis: a joint-carving theory of action might allow for the *exploitation* of another's agency in action without allowing another's actions *to be* one's actions.

[46] Note that in case (3), one's kicking and the other agent's action used as a means to kicking would likely be actions of different sorts. For instance, they presumably couldn't both be the sort of action which requires an exercise of a two-way power.

[47] Thanks to an anonymous referee for this objection.

such cases, one kicks by way of an external machine, but to trigger the machine, one need only intend to kick; thus, in this case, one kicks basically and by way of an external machine.

Second, I argue at length elsewhere that one can kick as a nonbasic action, for example, when one has a prosthetic leg that one controls via a handheld remote control (Kelley 2024). But here's a direct argument for that the meaning of the word 'kick' does not rule out kicking as a nonbasic action. Imagine you see Venita moving her leg in a kick-like manner, striking a soccer ball into a net. You'd describe Venita's leg movement as a kick, one which is an action of hers. If I then told you that she was controlling her leg movement using a handheld remote control, would you take back your description of her behavior as a kick? Would you say to yourself 'Oh, turns out Venita wasn't kicking after all'? Presumably, no. You'd instead feel you learned something about *how* Venita kicked, not about *whether* she kicked. So it appears that the meaning of the word 'kick' makes room for kicking as a nonbasic action.

It is also hard to see what the relevant difference would be between (2) and (3). Assuming the modal and explanatory conditions are met in both cases, what difference does it make whether the reliable external mechanism by which you kick is a machine or an agent? Perhaps the worry is that in case (3) as opposed to case (2), it could be that both you and something else control your movement; and perhaps a movement being your action relative to a description requires that you are the sole controller of it relative to that description. But this cannot be right, for it is often the case that one is not the sole controller of what are obviously one's actions: hosting a pot-luck is controlled in part by one's guests, mailing a letter cross-country is controlled in part by postal service workers, and executing an alley-oop is controlled in part by one's teammate. Perhaps the worry is instead that actions relative to bodily movement descriptions must be controlled solely by the mover. But again, this seems like an ad hoc restriction. For instance, an argument similar to the one just given would suggest that there is nothing in the meaning of the word 'kick' which rules out kicking by way of another agent's action. If you found out that the remote control that Venita uses to kick requires that someone approve every button she presses, you wouldn't conclude that Venita wasn't kicking after all; rather you'd feel like you learned even more about how she was kicking.

Of course, there are differences between (1) and (2) as well as between (2) and (3) that account for why all three kicks are actions *of different kinds*. The control theory of action offers tools for precisely identifying these differences, in fact: the kicks differ in the values they take along the three identified determination dimensions of action. The question is whether these differences justify distinguishing (1) from (2) in their statuses as actions and same with (2) and (3). And I've been trying to argue that the answer to this question is 'no'.

Perhaps, in one final attempt, one might appeal directly to intuitions about the applicability of the words

'action' and 'kick': (1) *seems* like a kick which is an action, and (2) and (3) don't. But intuitions are only decisive where there isn't some plausible alternative explanation for them. And indeed, it is easy to see why (1) seems like a kick which is an action while (2) or (3) don't: one has hardly any experience with kicks of the kind described in (2) and (3). Relative to how many kicks have been performed as actions in the way described in (1), there have been very few kicks performed as actions in the ways described in (2) and (3). That is, (2) and (3) describe highly unusual ways to perform a kick; this explains why one feels more hesitant about describing them as kicks and as actions. But unusual kicks are still kicks and unusual actions are still actions, just like unusual persons are still persons and unusual relationships are still relationships. A theory of action should leave room for diversity in how we do what we do. Thus, I conclude that there is no *relevant* difference between (1) and (3) that supports deeming (1) an action but not (3). We can indeed act by way of another's action.

6. Conclusion

I have offered a theory of action according to which action is reflexively controlled movement. I also offered a theory of reflexive control on which reflexive control requires meeting an aim condition, a modal condition, and an explanatory condition. But the control theory's contribution lies in its form as much as in its content. Indeed, seeing action as a determinable, the control theory keeps in plain sight the diversity within the category of action and thereby unifies a number of existing perspectives on agency. Moreover, by offering a schema for theories of particular forms of action, the control theory will hopefully facilitate investigation into ethically important kinds of action like intentional action and free action. And with a plausible theory of reflexive action, a very broad phenomenon associated with the category of action, we are closer to defending a plausible theory of action in the broadest sense. In these ways, the control theory of action is not meant to be the end of inquiry into the nature of action—instead, I suggest, it's the right place to start.

References

Aguilar, J. H. (2012), 'Basic Causal Deviance, Action Repertiores, and Reliability', *Philosophical Issues* **22**, 1–19.

Aguilar, J. H. and Buckareff, A. A. (2010), The Causal Theory of Action: Origins and Issues, *in* 'Causing Human Actions: New Perspectives on the Causal Theory of Action', The MIT Press.

Alvarez, M. (2009), 'Acting intentionally and acting for a reason', *Inquiry: An Interdisciplinary Journal of Philosophy* **52**(3), 293–305.

Alvarez, M. (2013), 'Agency and two-way powers', *Proceedings of the Aristotelian Society* **113**, 101–121.

Alvarez, M. and Hyman, J. (1998), 'Agents and their actions', *Philosophy* **73**(2), 219–245.

Anscombe, G. E. M. (1957), *Intention*, Harvard University Press.

Beddor, B. and Pavese, C. (2021), 'Practical Knowledge without Luminosity', *Mind* **131**(523), 917–934.

Bishop, J. (1989), *Natural Agency: An Essay on the Causal Theory of Action*, Cambridge University Press.

Bratman, M. (1987), *Intention, Plans, and Practical Reason*, Center for the Study of Language and Information.

Bratman, M. E. (2006), 'What is the accordion effect?', *The Journal of Ethics* **10**(1/2), 5–19.

Brozzo, C. (2017), 'Motor intentions: How intentions and motor representations come together', *Mind and Language* **32**(2), 231–256.

Burge, T. (2009), 'Primitive agency and natural norms', *Philosophy and Phenomenological Research* **79**(2), 251–278.

Camp, E. (2007), 'Thinking with maps', *Philosophical Perspectives* **21**(1), 145–182.

Chisholm, R. (1964), Human freedom and the self, *in* R. Kane, ed., 'Free Will', Blackwell.

Chisholm, R. (1985), Adverbs and subdeterminates, *in* E. Lepore and B. P. McLaughlin, eds, 'Actions and Events: Perspectives on the Philosophy of Donald Davidson', Blackwell.

Churchland, P. S. (1986), *Neurophilosophy: Toward a Unified Science of the Mind-Brain*, MIT Press.

Clark, A. and Chalmers, D. J. (1998), 'The extended mind', *Analysis* **58**(1), 7–19.

Davidson, D. (1980*a*), Agency, *in* 'Essays on Actions and Events', Clarendon Press.

Davidson, D. (1980*b*), Freedom to act, *in* 'Essays on Actions and Events', Clarendon Press.

Dennett, D. C. (1984), *Elbow Room: The Varieties of Free Will Worth Wanting*, MIT Press, London, England.

Fischer, J. M. and Ravizza, M. (1998), *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press.

Ford, A. (2011), Action and generality, *in* A. Ford, J. Hornsby and F. Stoutland, eds, 'Essays on Anscombe's Intention', Harvard University Press.

Ford, A. (2018), 'The province of human agency', *Noûs* **52**(3), 697–720.

Frankfurt, H. G. (1969), 'Alternate possibilities and moral responsibility', *The Journal of Philosophy* **66**(23), 829–839.

Frankfurt, H. G. (1971), 'Freedom of the will and the concept of a person', *The Journal of Philosophy* **68**(1), 5–20.

Frankfurt, H. G. (1978), 'The problem of action', *American Philosophical Quarterly* **15**(2), 157–162.

Funkhouser, E. (2006), 'The determinable-determinate relation', *Noûs* **40**(3), 548–569.

Háyek, A. (ms), '"If you had caught that plane, you would be dead by now": On Morgenbesser counterfactuals'.

Hyman, J. (2015), *Action, Knowledge, & Will*, Oxford University Press.

Jabr, F. (2022), 'The man who controls computers with his mind', *The New York Times* .
URL: *https://www.nytimes.com/2022/05/12/magazine/brain-computer-interface.html*

Kelley, M. (2024), 'How to perform a nonbasic action', *Noûs* **58**(1), 106–125.

Leitgeb, H. (2012), 'A Probabilistic Semantics for Counterfactuals. Part A', *Review of Symbolic Logic* **5**(1), 26–84.

Lewinsohn, J. (2024), 'The 'natural unintelligibility' of normative powers', *Jurisprudence* **15**(1), 5–34.

Lewis, D. (1973), 'Causation', *The Journal of Philosophy* **70**(17), 556–567.

Lewis, D. (2000), 'Causation as influence', *The Journal of Philosophy* **97**(4), 182–197.

Marcus, E. (2019), 'Reconciling practical knowledge with self-deception', *Mind* **128**(512), 1205–1225.

Milgram, E. (2010), Pluralism about action, *in* T. O'Connor and C. Sandis, eds, 'A Companion to the Philosophy of Action', Oxford: Wiley-Blackwell.

Milgram, E. (2022), Bounded agency, *in* L. Ferrero, ed., 'The Routledge Handbook of Philosophy of Agency', Routledge.

O'Brien, L. (2017), 'Actions as prime', *Royal Institute of Philosophy Supplement* **80**, 265–285.

O'Brien, L. (forthcoming), 'I, myself, move', *European Journal of Philosophy* .

Pacherie, E. (2011), 'Nonconceptual representations for action and the limits of intentional control', *Social Psychology* **42**(1), 67–73.

Pavese, C. (2021), 'Knowledge and mentality', *Wiley: Philosophical Perspectives* **35**(1), 359–382.

Peacocke, C. (1979), *Holistic Explanation*, Clarendon Press.

Prior, A. N. (1949), 'Determinables, determinates and determinants', *Mind* **58**(229), 1–20.

Rosenblueth, A., Wiener, N. and Bigelow, J. (1943), 'Behavior, purpose and teleology', *Philosophy of Science* **10**(1), 18–24.

Sartorio, C. (2010), Omissions and Causalism, *in* 'Causing Human Actions: New Perspectives on the Causal Theory of Action', The MIT Press.

Shepherd, J. (2014*a*), 'Causalism and intentional omission', *American Philosophical Quarterly* **51**(1), 15–26.

Shepherd, J. (2014*b*), 'The contours of control', *Philosophical Studies* **170**(3), 395–411. Publisher: Springer Verlag.

Shepherd, J. (2021), *The shape of agency: Control, action, skill, knowledge*, Oxford University Press.

Smith, M. (2021), 'Are actions bodily movements?', *Philosophical Explorations* **24**(3), 394–407.

Steward, H. (2012), *A Metaphysics for Freedom*, Oxford University Press, Oxford, GB.

Tenenbaum, S. (2007), *Appearances of the Good: An Essay on the Nature of Practical Reason*, Cambridge University Press.

Tomasello, M. (2022), *The Evolution of Agency: Behavioral Organization from Lizards to Humans*, The MIT Press.

Valaris, M. (2022), 'Knowledge out of control', *Philosophical Quarterly* **72**(3), 733–753.

Velleman, J. D. (1992), 'What happens when someone acts?', *Mind* **101**, 461–481.

Watson, G. (1975), 'Free agency', *Journal of Philosophy* **72**(April), 205–20.

Wilson, G. M. (1989), *The Intentionality of Human Action*, Stanford University Press.

Wilson, J. (2021), Determinables and Determinates, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', Spring 2021 edn, Metaphysics Research Lab, Stanford University.

Wu, W. (2016), 'Experts and Deviants: The Story of Agentive Control', *Philosophy and Phenomenological Research* **92**(2), 101–26.

Wu, W. (2023), *Movements of the Mind: A Theory of Attention, Intention and Action*, Oxford University Press, Oxford.